Category selectivity in human visual cortex

Peelen, Marius V.; Downing, Paul

Neuropsychologia

Category selectivity in human visual cortex: beyond visual object recognition

Marius V. Peelen[1] & Paul E. Downing[2]

1. Center for Mind/Brain Sciences (CIMeC), University of Trento, Italy

2. School of Psychology, Bangor University, UK

Keywords: category selectivity; ventral stream; object representation; scene perception

Address for correspondence:

Marius V. Peelen

Center for Mind/Brain Sciences (CIMeC)

University of Trento

Corso Bettini 31

Rovereto, TN 38068

Italy

Ph: +39 0464 808718

Fax: +39 0464 808690

Email: mariuspeelen@gmail.com

Abstract. Human ventral temporal cortex shows a categorical organization, with regions responding selectively to faces, bodies, tools, scenes, words, and other categories. Why is this? Traditional accounts explain category selectivity as arising within a hierarchical system dedicated to visual object recognition. For example, it has been proposed that category selectivity reflects the clustering of category-associated visual feature representations, or that it reflects category-specific computational algorithms needed to achieve view invariance. This visual object recognition framework has gained renewed interest with the success of deep neural network models trained to "recognize" objects: these hierarchical feed-forward networks show similarities to human visual cortex, including categorical separability. We argue that the object recognition framework is unlikely to fully account for category selectivity in visual cortex. Instead, we consider category selectivity in the context of other functions such as navigation, social cognition, tool use, and reading. Category-selective regions are activated during such tasks even in the absence of visual input and even in individuals with no prior visual experience. Further, they are engaged in close connections with broader domain-specific networks. Considering the diverse functions of these networks, category-selective regions likely encode their preferred stimuli in highly idiosyncratic formats; representations that are useful for navigation, social cognition, or reading are unlikely to be meaningfully similar to each other and to varying degrees may not be entirely visual. The demand for specific types of representations to support category-associated tasks may best account for category selectivity in visual cortex. This broader view invites new experimental and computational approaches.

Introduction

Evidence for category selectivity in the organization of human perception and cognition, and in the organization of high-level visual cortex, stretches back to the beginnings of modern cognitive neuroscience. Classic neuropsychological studies revealed striking dissociations in the performance of patients on judgments about living and non-living objects (Warrington & Shallice, 1984), as well as more selective deficits relating to specific categories such as human body parts (Sacchett & Humphreys, 1992). Inspired by these findings, functional neuroimaging studies went on to explore the topography and categorical nature of visual object representations in healthy participants. These revealed large-scale patterns of activity in high-level visual cortex that distinguish animate from inanimate categories (Bell, Hadj-Bouziane, Frihauf, Tootell, & Ungerleider, 2009; Kriegeskorte et al., 2008). Further, several focal regions exhibit strong and highly selective responses to more specific categories, such as scenes (Epstein, 2008), faces (Kanwisher & Yovel, 2006), bodies (Peelen & Downing, 2007), hands (Bracci, Ietswaart, Peelen, & Cavina-Pratesi, 2010), words (McCandliss, Cohen, & Dehaene, 2003), numerals (Shum et al., 2013), and tools (Martin, 2007). Numerous fMRI studies have established the regular and consistent arrangement of such regions across the ventral temporal cortex (VTC; Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Spiridon, Fischl, & Kanwisher, 2006).

While these findings have been widely replicated, debates continue over many basic questions about the apparently categorical organization of VTC: What computational benefits does it provide (Grill-Spector & Weiner, 2014; Leibo, Liao,

3

Anselmi, & Poggio, 2015)? To what extent does it reflect a response to evolutionary

pressures on brain organization (Caramazza & Shelton, 1998)? Can selectivity be

reduced to simpler principles such as visual feature preferences (Andrews, Watson, Rice,

& Hartley, 2015)? What do category-selective regions represent about their preferred

category (Gauthier, 2000; Haxby, Ishai, Chao, Ungerleider, & Martin, 2000)? How do

these regions connect to other parts of the brain (Hutchison, Culham, Everling,

Flanagan, & Gallivan, 2014; Osher et al., 2016)? How do they contribute to behavior

(Grill-Spector, Knouf, & Kanwisher, 2004; Pitcher, Charles, Devlin, Walsh, & Duchaine,

2009; Williams, Dang, & Kanwisher, 2007)? And how do they develop (Deen et al., 2017;

Dehaene et al., 2010; Grill-Spector, Golarai, & Gabrieli, 2008; Peelen, Glaser,

Vuilleumier, & Eliez, 2009; Saygin et al., 2016; Srihasam, Vincent, & Livingstone, 2014)?

The ventral temporal cortex as a visual object recognition system

Attempts to address these questions have generally adopted the view that VTC

constitutes the object recognition pathway (Goodale & Milner, 1992) – a pathway that

transforms low-level visual input into view-invariant object descriptors that permit the

perceptual identification of objects. That is, the primary goal of this pathway is to arrive

at a high-level description of object shape that can be compared against stored

representations in memory, and assessed (elsewhere in the brain) for emotional

meaning, reward significance and goal relevance (Kravitz, Saleem, Baker, Ungerleider, &

Mishkin, 2013). On this view, VTC constitutes the later stages of the object processing

hierarchy that begins with simple and complex cells in V1/V2 and extends to neurons

representing complex shape features with partial invariance to changes in viewpoint, size, and location (Haushofer, Livingstone, & Kanwisher, 2008; Op de Beeck, Torfs, & Wagemans, 2008; Riesenhuber & Poggio, 2000; Serre, Oliva, & Poggio, 2007; Tanaka, 1996).

This framework has deeply informed thinking about category selectivity. For example, the object-form topology account proposes that category-selective fMRI responses in VTC reflect the activation of visual shape representations that are mapped onto VTC in a continuous fashion (Haxby, et al., 2000; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Op de Beeck, Haushofer, & Kanwisher, 2008). This account can accommodate selectivity for categories that are characterized by shared shape features, such as faces (Caldara et al., 2006) or animals (Baldassi et al., 2013). Indeed, it has been argued that category preferences in VTC may be reduced to preferences for visual properties that are characteristic of those categories, even in the absence of category recognition. Such properties include spatial frequency (Rajimehr, Devaney, Bilenko, Young, & Tootell, 2011; Woodhead, Wise, Sereno, & Leech, 2011) and other low-level features (Rice, Watson, Hartley, & Andrews, 2014). For example, selectivity for scenes and buildings in the parahippocampal place area (PPA; Epstein & Kanwisher, 1998), located within the medial inanimate-preferring VTC, may reflect selectivity to cardinal orientations and right angles, features typical of man-made objects, buildings, and urban scenes (Nasr, Echavarria, & Tootell, 2014; Nasr & Tootell, 2012; but see Bryan, Julian, & Epstein, 2016).

An obvious following question, then, is why visual shape or feature representations of some categories, but not others, cluster together, giving rise to selectivity at the resolution of fMRI. Several proposals have been offered recently. One view is that category selectivity reflects the different computational algorithms that are brought to bear on the problem of achieving view-invariant object recognition for objects of different categories (Leibo, et al., 2015). Another proposal is that focal selectivity reflects the need for higher-order brain regions to efficiently read out category information (Grill-Spector & Weiner, 2014): clustering allows for the rapid categorization and recognition of stimuli at different hierarchical levels.

What these accounts have in common is that they build on the premise that the ventral visual cortex is a general-purpose system that has evolved to achieve visual object recognition by transforming simple visual features into high-level object descriptors. This approach has recently gained further popularity from the success of feedforward deep neural network (DNN) models (Krizhevsky, Sutskever, & Hinton, 2012) trained to successfully recognize objects (i.e., provide the correct object label) across a large set of diverse images, thus tackling the computational problem of transforming low-level visual input (pixels) to a view-invariant representation of object identity. Interestingly, these DNNs appear to capture aspects of visual cortex responses (Yamins & DiCarlo, 2016; but see Ullman, Assif, Fetaya, & Harari, 2016): the representational similarity structure of increasingly "deep" layers of a DNN showed a coarse correspondence to that of increasingly anterior regions in human visual cortex (Cichy, Khosla, Pantazis, Torralba, & Oliva, 2016). Furthermore, representations in deeper layers

of DNNs can be linearly combined so as to arrive at a categorical organization (e.g.,

animate-inanimate organization) that may resemble that observed in VTC (Khaligh-

Razavi & Kriegeskorte, 2014; Kubilius, Bracci, & Op de Beeck, 2016). Results like these

show that combinations of visual features can in principle give rise to a categorical

organization (Jozwik, Kriegeskorte, & Mur, 2016), in line with the view that category

selectivity arises as a function of visual object recognition constraints. DNN modeling

holds great promise for increasing our understanding of visual processing, and has

already shown success in explaining some aspects of visual cortex responses. Yet we

believe that feedforward models trained to identify objects from visual features will not

fully account for the categorical organization of VTC, for the reasons discussed in the

following sections.

The ventral temporal cortex: beyond visual object recognition

While the VTC is typically studied and modeled in the context of visual object

recognition and categorization, clearly these are not the only tasks our visual system

must perform (Figure 1). Incoming visual input supports diverse tasks such as spatial

navigation, inferring the emotions of other people, and interpreting written text. These

tasks each require tailored representations of the outside world (Cox, 2014): the types

of representation that are useful for navigation are fundamentally different from those

that are useful for emotion recognition. We argue that the demand for these diverse

types of representations – needed to efficiently perform the range of tasks humans face

every day – is a driving force behind the evolution and development of category selectivity in VTC.

On this view, category-selective regions in VTC represent the world in a way that is relevant for solving tasks associated specifically with that category. For example, scene-selective regions represent aspects of spatial layout that are relevant for navigation (Epstein, 2008); body-selective regions represent body configurations that are informative about others' actions and social characteristics (Downing & Peelen, 2011); face-selective regions represent face parts and their configuration to support person identification (Freiwald, Duchaine, & Yovel, 2016); and tool-selective regions represent tools in terms of the hand postures and movements associated with their use (Beauchamp, Lee, Haxby, & Martin, 2002; Bracci & Peelen, 2013; Perini, Caramazza, & Peelen, 2014).

As such, category-selective regions in VTC should be considered as integral and interactive parts of large-scale brain networks dedicated to processes such as reading, navigation, social cognition, and tool use (for related views, see: Mahon & Caramazza, 2011; Martin, 2007; Price & Devlin, 2011; Reich, Szwed, Cohen, & Amedi, 2011). Indeed, what these regions represent about their preferred category can only be understood by considering the goals that must be achieved by the whole systems in which they reside.

Category selectivity can be dissociated from visual object processing

This view is in line with findings of studies showing that category-selective activity in VTC can be dissociated from visual object processing. That is, rather than simply

reflecting a high-level representation of the visual input, category-selective activity

appears to be closely tied to the interpretation of an object (e.g., a tool) as belonging to

that category and the cognition (or behavior) associated with that category (e.g., tool

use). For example, briefly training participants to use a set of complex novel objects as

tools results in activity in the tool-selective left lateral occipitotemporal cortex, together

with regions in parietal and premotor cortex implicated in tool use, when those objects

are later passively viewed (Weisberg, van Turennout, & Martin, 2007). In other words,

objects evoke tool-selective responses only when observers know, through experience,

how these objects can be used as tools (Figure 2a). Similarly, judging the type of action

that a tool object uses (such as twist or squeeze) greatly amplifies the responses of the

left occipitotemporal tool-selective region, relative to an equally demanding judgment

of the typical location in which the object is found (Perini, et al., 2014). Because visual

object recognition is equated across conditions, the increase in activity must reflect

activity that is specifically relevant for the tool use system.

Other evidence comes from studies showing that activity in the right fusiform

gyrus — a region broadly selective for faces, bodies, and animals — can be evoked by

simple geometric shapes that, through their movements, are interpreted as social

agents (Castelli, Happe, Frith, & Frith, 2000; Gobbini, Koralek, Bryan, Montgomery, &

Haxby, 2007; Martin & Weisberg, 2003; Schultz et al., 2003). Again, these studies

dissociate visual object recognition (the visual processing of the simple shapes) from

category-selective responses. Finally, recent studies have experimentally dissociated

visual shape-selective from category-selective responses using closely matched stimulus

pairs (e.g., snake-rope), showing category selectivity in VTC after accounting for the contributions of visual shape features (Figure 2b; Bracci & Op de Beeck, 2016; Bryan, et al., 2016; Kaiser, Azzalini, & Peelen, 2016; Macdonald & Culham, 2015; Proklova, Kaiser, & Peelen, 2016).

Altogether, these studies show that the processing of visual features is not sufficient to account for category-selective responses in occipitotemporal cortex. Rather, they show that category-selective responses are closely aligned with knowledge of what a thing means to the observer; that is, the cognitions and behaviors an object is associated with.

Category selectivity in congenitally blind individuals

Further evidence that cannot be easily accommodated by visual object recognition accounts of category selectivity comes from studies investigating category-selective responses in individuals who have been blind from, or shortly after, birth. Even though these individuals do not have visual representations, the category selectivity for objects presented in non-visual modalities in this group is still observed, at least for some categories (for reviews, see Bi, Wang, & Caramazza, 2016; Ricciardi, Bonino, Pellegrini, & Pietrini, 2014). For example, when blind participants listened to words describing landmark objects or tools, selective activity similar to that found in sighted individuals was revealed in VTC (Figure 3a; He et al., 2013; Mahon, Anzellotti, Schwarzbach, Zampini, & Caramazza, 2009; Peelen et al., 2013). Selective activity in the PPA was also observed when congenitally blind participants haptically explored miniature scenes,

relative to complexity-matched objects (Wolbers, Klatzky, Loomis, Wutte, & Giudice, 2011). Similarly, haptic body-perception induced activity in the body-selective EBA in blind participants (Kitada et al., 2014). Other studies have shown focal activity in the visual word form area (VWFA) when blind participants read Braille (Buchel, Price, & Friston, 1998; Reich, et al., 2011). Finally, using a sensory-substitution device (SSD) that transforms images into auditory "soundscapes", blind participants showed activity in VWFA when using the SSD to read words (Figure 3b; Striem-Amit, Cohen, Dehaene, & Amedi, 2012) and in EBA when "viewing" bodies (Striem-Amit & Amedi, 2014).

Importantly, the degree to which category selectivity depends on visual input and visual experience differs across categories and across anatomical regions of VTC. For example, the study that provided evidence for landmark object selectivity in the PPA of blind individuals showed a clear absence of selectivity for animals in the posterior fusiform gyrus of the same blind group (He, et al., 2013). A recent study directly compared the connectional and functional profiles (or "fingerprints") of VTC voxels across blind and sighted individuals, providing maps of the degree to which regions were similar or different across groups in terms of their categorical response profiles and whole-brain resting-state connectivity patterns (Wang et al., 2015). Interestingly, both functional and connectional maps showed regions of VTC that were virtually indistinguishable across groups (including tool-, body-, and scene-selective regions) as well as regions that responded and connected very differently as a function of visual experience (including animate -selective regions in posterior fusiform gyrus).

These findings provide information about the types of representations in these regions. For example, they suggest that representations in posterior fusiform gyrus may reflect properties of animals (including human faces and bodies) that are primarily sensed through the visual modality and that may thus be lacking in blind individuals. By contrast, representations in other parts of VTC may reflect properties that are not exclusively visual. More research is needed to establish what exactly these non-visual representations are. One possibility is that, in the absence of visual input, category-selective regions in visual cortex start to process input from other modalities such as audition and touch (Pascual-Leone & Hamilton, 2001). Alternatively, and more in line with the finding of similar functional connectivity in blind and sighted (Wang et al., 2015), these regions may encode object properties in an amodal format in both blind and sighted individuals, for example of spatial layout, body posture, or object shape (Peelen et al., 2014). While vision is typically the most reliable source for knowing about these properties, congenitally blind individuals also acquire (and thus represent) this knowledge.

More generally, evidence that specific visual features, visual input and even visual experience are not necessary to elicit the majority of category-selective responses in VTC demonstrates that this organization as a whole does not principally develop to support visual object recognition. Rather than reflecting constraints within a feedforward object recognition system, category selectivity may be better understood in the context of broader domain-specific networks (Bi, et al., 2016; Mahon & Caramazza, 2011).

Category-selective regions as integral parts of domain-specific networks

Our view puts category-selective regions closely within broader domain-specific networks involved in performing daily-life tasks. For these regions to successfully support functions like reading, navigation, tool use, social cognition or action understanding, they must interact closely with brain-wide networks involved in these functions, such as regions representing heading direction (Furlan, Wann, & Smith, 2014), tool use (Lewis, 2006), or others' mental states (Koster-Hale, Bedny, & Saxe, 2014). There is increasing empirical evidence for selective connectivity between category-selective regions in VTC and regions in other parts of the brain serving related functions (Bracci, Cavina-Pratesi, Ietswaart, Caramazza, & Peelen, 2012; Hutchison, et al., 2014; Simmons & Martin, 2012). Furthermore, these connectivity patterns accurately predict category selectivity in VTC (Osher, et al., 2016; Saygin et al., 2011; Saygin, et al., 2016; Wang, et al., 2015). Considering the anatomical location of these regions, the specific role they play within their broader networks is likely related to more perceptual aspects, in that they represent information that is typically (but not exclusively; see previous section) provided by the visual modality.

What accounts for category selectivity in visual cortex?

Marr (1982) famously noted that we need to understand what the goal of a computation is before asking what types of representations are involved and how these are implemented. We argue that category-selective regions in VTC exist because they

contribute to wider networks that serve critical behavioral goals such as navigation, recognizing conspecifics, using tools, and understanding others' actions and emotions. Situated as they are in the ventral stream, these regions collectively benefit from an architecture that allows them to encode critical stimuli in robust representations that are tolerant to functionally irrelevant variation. As such, there are undoubtedly similarities in the computational principles underlying the processing of objects of different categories (Taylor & Downing, 2011). At the same time, however, to support their unique functional roles, these different regions must encode their preferred stimuli in highly idiosyncratic formats. Many of the task-critical dimensions encoded in tool, word, number, face, body, and scene representations (or neural "spaces") are unlikely to be meaningfully similar to each other, and these dimensions vary in the degree to which they are exclusively visual. For example: in calculation and in reading, numerals and letters must be identified across variations in fonts and styles; for navigation, boundaries and edges are important, where textures and colors are much less so; a person remains the same person regardless of limb posture and movement, yet those features are critical for action understanding; and useful tool representations must capture the tool's shape as it relates to hand actions, including a distinction between the proximal and distal (effector) ends of the object. Numerals do not have effectors; hammers do not come in different fonts; scenes do not hold different postures. Directly comparing neural representations of these natural kinds without taking these distinctions into account is a comparison of apples to oranges.

A full account of category selectivity in VTC, we argue, should encompass these varied properties, which arise from the unique functional pressures imposed by different real-world tasks, and which do not naturally arise from a view of selective regions as being a family of similar nodes that are situated together at the apex of a homogenous visual feature-processing hierarchy.

Future directions

A corollary of the present view is that we will be better placed to study and model category selectivity in the context of naturalistic tasks. Although studies measuring responses to highly reduced and artificial stimuli will surely continue to be valuable, a full understanding of category selectivity in VTC requires consideration and approximation of the conditions that prompted the evolution and development of category selectivity in the first place. We expect that future progress will involve the use of more realistic environments (e.g., Hasson, Malach, & Heeger, 2010; Snow et al., 2011) to investigate navigation (e.g., Julian, Ryan, Hamilton, & Epstein, 2016), reading (e.g., Schuster, Hawelka, Richlan, Ludersdorfer, & Hutzler, 2015), object detection (e.g., Peelen & Kastner, 2014), tool use (e.g., Gallivan, McLean, Valyear, & Culham, 2013), and social cognition (e.g., Hasson & Frith, 2016). Finally, computational approaches may provide further support for our view by training DNNs on tasks other than object recognition. It is only in the context of these tasks that we will arrive at a complete picture of how category-selective activity in ventral temporal cortex develops and how it contributes to behavior.

References

Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of

natural images predict topographic patterns of neural response in the ventral

visual pathway. *J Vis, 15*(7), 3. doi: 10.1167/15.7.3

2296936 [pii]

Baldassi, C., Alemi-Neissi, A., Pagan, M., Dicarlo, J. J., Zecchina, R., & Zoccolan, D. (2013).

Shape similarity, better than semantic membership, accounts for the structure of

visual object representations in a population of monkey inferotemporal neurons.

PLoS Comput Biol, 9(8), e1003167. doi: 10.1371/journal.pcbi.1003167

PCOMPBIOL-D-13-00131 [pii]

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion

processing streams for manipulable objects and human movements. *Neuron,

34*(1), 149-159. doi: S0896627302006426 [pii]

Bell, A. H., Hadj-Bouziane, F., Frihauf, J. B., Tootell, R. B., & Ungerleider, L. G. (2009).

Object representations in the temporal cortex of monkeys and humans as

revealed by functional magnetic resonance imaging. *J Neurophysiol, 101*(2), 688-

700. doi: 10.1152/jn.90657.2008

90657.2008 [pii]

Bi, Y., Wang, X., & Caramazza, A. (2016). Object Domain and Modality in the Ventral

Visual Pathway. *Trends Cogn Sci, 20*(4), 282-290. doi: 10.1016/j.tics.2016.02.002

S1364-6613(16)00043-7 [pii]

Bracci, S., Cavina-Pratesi, C., Ietswaart, M., Caramazza, A., & Peelen, M. V. (2012).

Closely overlapping responses to tools and hands in left lateral occipitotemporal

cortex. *J Neurophysiol, 107*(5), 1443-1456. doi: jn.00619.2011 [pii]

10.1152/jn.00619.2011

Bracci, S., Ietswaart, M., Peelen, M. V., & Cavina-Pratesi, C. (2010). Dissociable neural

responses to hands and non-hand body parts in human left extrastriate visual

cortex. *J Neurophysiol, 103*(6), 3389-3397. doi: jn.00215.2010 [pii]

10.1152/jn.00215.2010

Bracci, S., & Op de Beeck, H. (2016). Dissociations and Associations between Shape and

Category Representations in the Two Visual Pathways. *J Neurosci, 36*(2), 432-

444. doi: 10.1523/JNEUROSCI.2314-15.2016

36/2/432 [pii]

Bracci, S., & Peelen, M. V. (2013). Body and object effectors: the organization of object

representations in high-level visual cortex reflects body-object interactions. *J

Neurosci, 33*(46), 18247-18258. doi: 10.1523/JNEUROSCI.1322-13.2013

33/46/18247 [pii]

Bryan, P. B., Julian, J. B., & Epstein, R. A. (2016). Rectilinear Edge Selectivity Is

Insufficient to Explain the Category Selectivity of the Parahippocampal Place

Area. *Front Hum Neurosci, 10*, 137. doi: 10.3389/fnhum.2016.00137

Buchel, C., Price, C., & Friston, K. (1998). A multimodal language region in the ventral

visual pathway. *Nature, 394*(6690), 274-277. doi: 10.1038/28389

Caldara, R., Seghier, M. L., Rossion, B., Lazeyras, F., Michel, C., & Hauert, C. A. (2006). The fusiform face area is tuned for curvilinear patterns with more high-contrasted elements in the upper part. *Neuroimage, 31*(1), 313-319. doi: S1053-8119(05)02537-1 [pii]

10.1016/j.neuroimage.2005.12.011

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain the animate-inanimate distinction. *J Cogn Neurosci, 10*(1), 1-34.

Castelli, F., Happe, F., Frith, U., & Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage, 12*(3), 314-325. doi: 10.1006/nimg.2000.0612

S1053-8119(00)90612-8 [pii]

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci Rep, 6*, 27755. doi: 10.1038/srep27755

srep27755 [pii]

Cox, D. D. (2014). Do we understand high-level vision? *Curr Opin Neurobiol, 25*, 187-193. doi: 10.1016/j.conb.2014.01.016

S0959-4388(14)00031-2 [pii]

Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., . . . Saxe, R. (2017). Organization of high-level visual cortex in human infants. *Nature Communications, 8*. doi: Artn 13995

10.1038/Ncomms13995

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., . . . Cohen,

L. (2010). How learning to read changes the cortical networks for vision and

language. *Science, 330*(6009), 1359-1364. doi: 10.1126/science.1194140

science.1194140 [pii]

Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006).

Domain specificity in visual cortex. *Cereb Cortex, 16*(10), 1453-1461. doi: bhj086

[pii]

10.1093/cercor/bhj086

Downing, P. E., & Peelen, M. V. (2011). The role of occipitotemporal body-selective

regions in person perception. *Cogn Neurosci, 2*(3-4), 186-203. doi:

10.1080/17588928.2011.582945

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual

environment. *Nature, 392*(6676), 598-601. doi: 10.1038/33402

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial

navigation. *Trends Cogn Sci, 12*(10), 388-396. doi: 10.1016/j.tics.2008.07.004

S1364-6613(08)00199-X [pii]

Freiwald, W., Duchaine, B., & Yovel, G. (2016). Face Processing Systems: From Neurons

to Real-World Social Perception. *Annu Rev Neurosci, 39*, 325-346. doi:

10.1146/annurev-neuro-070815-013934

Furlan, M., Wann, J. P., & Smith, A. T. (2014). A representation of changing heading

direction in human cortical areas pVIP and CSv. *Cereb Cortex, 24*(11), 2848-2858.

doi: 10.1093/cercor/bht132

bht132 [pii]

Gallivan, J. P., McLean, D. A., Valyear, K. F., & Culham, J. C. (2013). Decoding the neural

mechanisms of human tool use. *Elife, 2*, e00425. doi: 10.7554/eLife.00425

00425 [pii]

Gauthier, I. I. (2000). What constrains the organization of the ventral temporal cortex?

Trends Cogn Sci, 4(1), 1-2. doi: S1364-6613(99)01416-3 [pii]

Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two

takes on the social brain: a comparison of theory of mind tasks. *J Cogn Neurosci,*

19(11), 1803-1814. doi: 10.1162/jocn.2007.19.11.1803

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and

action. *Trends Neurosci, 15*(1), 20-25. doi: 0166-2236(92)90344-8 [pii]

Grill-Spector, K., Golarai, G., & Gabrieli, J. (2008). Developmental neuroimaging of the

human ventral visual cortex. *Trends Cogn Sci, 12*(4), 152-162. doi:

10.1016/j.tics.2008.01.009

S1364-6613(08)00057-0 [pii]

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves

face perception, not generic within-category identification. *Nature Neuroscience,*

7(5), 555-562. doi: 10.1038/nn1224

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral

temporal cortex and its role in categorization. *Nat Rev Neurosci, 15*(8), 536-548.

doi: 10.1038/nrn3747

nrn3747 [pii]

Hasson, U., & Frith, C. D. (2016). Mirroring and beyond: coupled dynamics as a

generalized framework for modelling social interactions. *Philos Trans R Soc Lond

B Biol Sci, 371*(1693). doi: 10.1098/rstb.2015.0366

20150366 [pii]

rstb.2015.0366 [pii]

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during

natural stimulation. *Trends Cogn Sci, 14*(1), 40-48. doi:

10.1016/j.tics.2009.10.011

S1364-6613(09)00239-3 [pii]

Haushofer, J., Livingstone, M. S., & Kanwisher, N. (2008). Multivariate patterns in object-

selective cortex dissociate perceptual and physical shape similarity. *PLoS Biol,

6*(7), e187. doi: 10.1371/journal.pbio.0060187

08-PLBI-RA-0602 [pii]

Haxby, J. V., Ishai, I. I., Chao, L. L., Ungerleider, L. G., & Martin, I. I. (2000). Object-form

topology in the ventral temporal lobeResponse to I. Gauthier (2000). *Trends

Cogn Sci, 4*(1), 3-4. doi: S1364661399014230 [pii]

He, C., Peelen, M. V., Han, Z., Lin, N., Caramazza, A., & Bi, Y. (2013). Selectivity for large

nonmanipulable objects in scene-selective visual cortex does not require visual

experience. *Neuroimage, 79*, 1-9. doi: 10.1016/j.neuroimage.2013.04.051

S1053-8119(13)00398-4 [pii]

Hutchison, R. M., Culham, J. C., Everling, S., Flanagan, J. R., & Gallivan, J. P. (2014).

Distinct and distributed functional connectivity patterns across cortex reflect the

domain-specific constraints of object, face, scene, body, and tool category-

selective modules in the ventral visual pathway. *Neuroimage, 96*, 216-236. doi:

10.1016/j.neuroimage.2014.03.068

S1053-8119(14)00229-8 [pii]

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed

representation of objects in the human ventral visual pathway. *Proc Natl Acad

Sci U S A, 96*(16), 9379-9384.

Jozwik, K. M., Kriegeskorte, N., & Mur, M. (2016). Visual features as stepping stones

toward semantics: Explaining object similarity in IT and perception with non-

negative least squares. *Neuropsychologia, 83*, 201-226. doi:

10.1016/j.neuropsychologia.2015.10.023

S0028-3932(15)30199-8 [pii]

Julian, J. B., Ryan, J., Hamilton, R. H., & Epstein, R. A. (2016). The Occipital Place Area Is

Causally Involved in Representing Environmental Boundaries during Navigation.

Curr Biol, 26(8), 1104-1109. doi: 10.1016/j.cub.2016.02.066

S0960-9822(16)30177-4 [pii]

Kaiser, D., Azzalini, D. C., & Peelen, M. V. (2016). Shape-independent object category

responses revealed by MEG and fMRI decoding. *J Neurophysiol, 115*(4), 2246-

2250. doi: 10.1152/jn.01074.2015

jn.01074.2015 [pii]

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized

for the perception of faces. *Philos Trans R Soc Lond B Biol Sci, 361*(1476), 2109-

2128. doi: X83871K50X365124 [pii]

10.1098/rstb.2006.1934

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised,

models may explain IT cortical representation. *PLoS Comput Biol, 10*(11),

e1003915. doi: 10.1371/journal.pcbi.1003915

PCOMPBIOL-D-14-00531 [pii]

Kitada, R., Yoshihara, K., Sasaki, A. T., Hashiguchi, M., Kochiyama, T., & Sadato, N.

(2014). The brain network underlying the recognition of hand gestures in the

blind: the supramodal role of the extrastriate body area. *J Neurosci, 34*(30),

10096-10108. doi: 10.1523/JNEUROSCI.0500-14.2014

34/30/10096 [pii]

Koster-Hale, J., Bedny, M., & Saxe, R. (2014). Thinking about seeing: perceptual sources

of knowledge are encoded in the theory of mind brain regions of sighted and

blind adults. *Cognition, 133*(1), 65-78. doi: 10.1016/j.cognition.2014.04.006

S0010-0277(14)00067-5 [pii]

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The

ventral visual pathway: an expanded neural framework for the processing of

object quality. *Trends Cogn Sci, 17*(1), 26-49. doi: 10.1016/j.tics.2012.10.011

S1364-6613(12)00247-1 [pii]

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., . . . Bandettini, P.

A. (2008). Matching categorical object representations in inferior temporal

cortex of man and monkey. *Neuron, 60*(6), 1126-1141. doi:

10.1016/j.neuron.2008.10.043

S0896-6273(08)00943-4 [pii]

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep

Convolutional Neural Networks. *Advances in Neural Information Processing

Systems, 25*, 1097-1105.

Kubilius, J., Bracci, S., & Op de Beeck, H. P. (2016). Deep Neural Networks as a

Computational Model for Human Shape Sensitivity. *PLoS Comput Biol, 12*(4),

e1004896. doi: 10.1371/journal.pcbi.1004896

PCOMPBIOL-D-15-01441 [pii]

Leibo, J. Z., Liao, Q., Anselmi, F., & Poggio, T. (2015). The Invariance Hypothesis Implies

Domain-Specific Regions in Visual Cortex. *PLoS Comput Biol, 11*(10), e1004390.

doi: 10.1371/journal.pcbi.1004390

PCOMPBIOL-D-14-01081 [pii]

Lewis, J. W. (2006). Cortical networks related to human use of tools. *Neuroscientist,

12*(3), 211-231. doi: 12/3/211 [pii]

10.1177/1073858406288327

Macdonald, S. N., & Culham, J. C. (2015). Do human brain areas involved in visuomotor
actions show a preference for real tools over visually similar non-tools?
Neuropsychologia, 77, 35-41. doi: 10.1016/j.neuropsychologia.2015.08.004
S0028-3932(15)30120-2 [pii]

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009).
Category-specific organization in the human brain does not require visual
experience. *Neuron, 63*(3), 397-405. doi: S0896-6273(09)00541-8 [pii]
10.1016/j.neuron.2009.07.012

Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge
in the brain? *Trends Cogn Sci, 15*(3), 97-103. doi: S1364-6613(11)00005-2 [pii]
10.1016/j.tics.2011.01.004

Marr, D. (1982). *Vision: A Computational Approach*: Freeman & Co.

Martin, A. (2007). The representation of object concepts in the brain. *Annu Rev Psychol,
58*, 25-45. doi: 10.1146/annurev.psych.57.102904.190143

Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and
mechanical concepts. *Cogn Neuropsychol, 20*(3-6), 575-587. doi:
10.1080/02643290342000005

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: expertise
for reading in the fusiform gyrus. *Trends Cogn Sci, 7*(7), 293-299. doi:
S1364661303001347 [pii]

Nasr, S., Echavarria, C. E., & Tootell, R. B. (2014). Thinking outside the box: rectilinear

shapes selectively activate scene-selective cortex. *J Neurosci, 34*(20), 6721-6735.

doi: 10.1523/JNEUROSCI.4802-13.2014

JNEUROSCI.4802-13.2014 [pii]

Nasr, S., & Tootell, R. B. (2012). A cardinal orientation bias in scene-selective visual

cortex. *J Neurosci, 32*(43), 14921-14926. doi: 10.1523/JNEUROSCI.2036-12.2012

32/43/14921 [pii]

Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data:

maps, modules and dimensions. *Nat Rev Neurosci, 9*(2), 123-135. doi:

10.1038/nrn2314

nrn2314 [pii]

Op de Beeck, H. P., Torfs, K., & Wagemans, J. (2008). Perceived shape similarity among

unfamiliar objects and the organization of the human object vision pathway. *J

Neurosci, 28*(40), 10111-10123. doi: 10.1523/JNEUROSCI.2511-08.2008

28/40/10111 [pii]

Osher, D. E., Saxe, R. R., Koldewyn, K., Gabrieli, J. D., Kanwisher, N., & Saygin, Z. M.

(2016). Structural Connectivity Fingerprints Predict Cortical Selectivity for

Multiple Visual Categories across Cortex. *Cereb Cortex, 26*(4), 1668-1683. doi:

10.1093/cercor/bhu303

bhu303 [pii]

Pascual-Leone, A., & Hamilton, R. (2001). The metamodal organization of the brain. *Prog

Brain Res, 134*, 427-445.

Peelen, M. V., Bracci, S., Lu, X., He, C., Caramazza, A., & Bi, Y. (2013). Tool selectivity in

left occipitotemporal cortex develops without vision. *J Cogn Neurosci*.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nat

Rev Neurosci, 8*(8), 636-648. doi: nrn2195 [pii]

10.1038/nrn2195

Peelen, M. V., Glaser, B., Vuilleumier, P., & Eliez, S. (2009). Differential development of

selectivity for faces and bodies in the fusiform gyrus. *Dev Sci, 12*(6), F16-25. doi:

10.1111/j.1467-7687.2009.00916.x

DESC916 [pii]

Peelen, M.V., He, C., Han, Z., Caramazza, A., & Bi, Y. (2014). Nonvisual and visual object

shape representations in occipitotemporal cortex: Evidence from congenitally

blind and sighted adults. *J Neurosci, 34*(1), 163-170.

Peelen, M. V., & Kastner, S. (2014). Attention in the real world: toward understanding its

neural basis. *Trends Cogn Sci, 18*(5), 242-250. doi: 10.1016/j.tics.2014.02.004

S1364-6613(14)00047-3 [pii]

Perini, F., Caramazza, A., & Peelen, M. V. (2014). Left occipitotemporal cortex

contributes to the discrimination of tool-associated hand actions: fMRI and TMS

evidence. *Front Hum Neurosci, 8*, 591. doi: 10.3389/fnhum.2014.00591

Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., & Duchaine, B. (2009). Triple dissociation

of faces, bodies, and objects in extrastriate cortex. *Curr Biol, 19*(4), 319-324. doi:

S0960-9822(09)00543-0 [pii]

10.1016/j.cub.2009.01.007

Price, C. J., & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal

   contributions to reading. *Trends Cogn Sci, 15*(6), 246-253. doi:

   10.1016/j.tics.2011.04.001

S1364-6613(11)00057-X [pii]

Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling Representations of Object

   Shape and Object Category in Human Visual Cortex: The Animate-Inanimate

   Distinction. *J Cogn Neurosci, 28*(5), 680-692. doi: 10.1162/jocn_a_00924

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. (2011). The

   "parahippocampal place area" responds preferentially to high spatial frequencies

   in humans and monkeys. *PLoS Biol, 9*(4), e1000608. doi:

   10.1371/journal.pbio.1000608

Reich, L., Szwed, M., Cohen, L., & Amedi, A. (2011). A ventral visual stream reading

   center independent of visual experience. *Curr Biol, 21*(5), 363-368. doi: S0960-

   9822(11)00063-7 [pii]

10.1016/j.cub.2011.01.040

Ricciardi, E., Bonino, D., Pellegrini, S., & Pietrini, P. (2014). Mind the blind brain to

   understand the sighted one! Is there a supramodal cortical functional

   architecture? *Neurosci Biobehav Rev, 41*, 64-77. doi:

   10.1016/j.neubiorev.2013.10.006

S0149-7634(13)00230-3 [pii]

Rice, G. E., Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Low-level image

   properties of visual objects predict patterns of neural response across category-

selective regions of the ventral visual pathway. *J Neurosci, 34*(26), 8837-8844.

doi: 10.1523/JNEUROSCI.5265-13.2014

34/26/8837 [pii]

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nat Neurosci, 3 Suppl*, 1199-1204. doi: 10.1038/81479

Sacchett, C., & Humphreys, G. W. (1992). Calling a Squirrel a Squirrel but a Canoe a Wigwam - a Category-Specific Deficit for Artifactual Objects and Body Parts. *Cognitive Neuropsychology, 9*(1), 73-86. doi: Doi 10.1080/02643299208252053

Saygin, Z. M., Osher, D. E., Koldewyn, K., Reynolds, G., Gabrieli, J. D., & Saxe, R. R. (2011). Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nat Neurosci, 15*(2), 321-327. doi: 10.1038/nn.3001

nn.3001 [pii]

Saygin, Z. M., Osher, D. E., Norton, E. S., Youssoufian, D. A., Beach, S. D., Feather, J., . . . Kanwisher, N. (2016). Connectivity precedes function in the development of the visual word form area. *Nat Neurosci, 19*(9), 1250-1255. doi: 10.1038/nn.4354

nn.4354 [pii]

Schultz, R. T., Grelotti, D. J., Klin, A., Kleinman, J., Van der Gaag, C., Marois, R., & Skudlarski, P. (2003). The role of the fusiform face area in social cognition: implications for the pathobiology of autism. *Philos Trans R Soc Lond B Biol Sci, 358*(1430), 415-427. doi: 10.1098/rstb.2002.1208

Schuster, S., Hawelka, S., Richlan, F., Ludersdorfer, P., & Hutzler, F. (2015). Eyes on words: A fixation-related fMRI study of the left occipito-temporal cortex during

self-paced silent reading of words and pseudowords. *Sci Rep, 5*, 12686. doi:

10.1038/srep12686

srep12686 [pii]

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid

categorization. *Proc Natl Acad Sci U S A, 104*(15), 6424-6429. doi: 0700622104

[pii]

10.1073/pnas.0700622104

Shum, J., Hermes, D., Foster, B. L., Dastjerdi, M., Rangarajan, V., Winawer, J., . . . Parvizi,

J. (2013). A brain area for visual numerals. *J Neurosci, 33*(16), 6709-6715. doi:

10.1523/JNEUROSCI.4558-12.2013

33/16/6709 [pii]

Simmons, W. K., & Martin, A. (2012). Spontaneous resting-state BOLD fluctuations

reveal persistent domain-specific neural networks. *Soc Cogn Affect Neurosci,*

7(4), 467-475. doi: nsr018 [pii]

10.1093/scan/nsr018

Snow, J. C., Pettypiece, C. E., McAdam, T. D., McLean, A. D., Stroman, P. W., Goodale, M.

A., & Culham, J. C. (2011). Bringing the real world into the fMRI scanner:

repetition effects for pictures versus real objects. *Sci Rep, 1*, 130. doi:

10.1038/srep00130

Spiridon, M., Fischl, B., & Kanwisher, N. (2006). Location and spatial profile of category-

specific regions in human extrastriate cortex. *Hum Brain Mapp, 27*(1), 77-89. doi:

10.1002/hbm.20169

Srihasam, K., Vincent, J. L., & Livingstone, M. S. (2014). Novel domain formation reveals

proto-architecture in inferotemporal cortex. *Nature Neuroscience, 17*(12), 1776-

1783. doi: 10.1038/nn.3855

Striem-Amit, E., & Amedi, A. (2014). Visual cortex extrastriate body-selective area

activation in congenitally blind people "seeing" by using sounds. *Curr Biol, 24*(6),

687-692. doi: 10.1016/j.cub.2014.02.010

S0960-9822(14)00148-1 [pii]

Striem-Amit, E., Cohen, L., Dehaene, S., & Amedi, A. (2012). Reading with sounds:

sensory substitution selectively activates the visual word form area in the blind.

Neuron, 76(3), 640-652. doi: 10.1016/j.neuron.2012.08.026

S0896-6273(12)00763-5 [pii]

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu Rev Neurosci, 19*, 109-

139. doi: 10.1146/annurev.ne.19.030196.000545

Taylor, J.C., & Downing, P.E. (2011). Division of labor between lateral and ventral

extrastriate representations of faces, bodies, and objects. *J Cogn Neurosci,

23*(12), 4122-4137. doi: 10.1162/jocn_a_00091

Ullman, S., Assif, L., Fetaya, E., & Harari, D. (2016). Atoms of recognition in human and

computer vision. *Proc Natl Acad Sci U S A, 113*(10), 2744-2749. doi:

10.1073/pnas.1513198113

1513198113 [pii]

Wang, X., Peelen, M. V., Han, Z., He, C., Caramazza, A., & Bi, Y. (2015). How Visual Is the

Visual Cortex? Comparing Connectional and Functional Fingerprints between

Congenitally Blind and Sighted Individuals. *J Neurosci, 35*(36), 12545-12559. doi:

10.1523/JNEUROSCI.3914-14.2015

35/36/12545 [pii]

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain,*

107 (Pt 3), 829-854.

Weisberg, J., van Turennout, M., & Martin, A. (2007). A neural system for learning about

object function. *Cereb Cortex, 17*(3), 513-521. doi: bhj176 [pii]

10.1093/cercor/bhj176

Williams, M. A., Dang, S., & Kanwisher, N. G. (2007). Only some spatial patterns of fMRI

response are read out in task performance. *Nature Neuroscience, 10*(6), 685-686.

doi: 10.1038/nn1900

Wolbers, T., Klatzky, R. L., Loomis, J. M., Wutte, M. G., & Giudice, N. A. (2011). Modality-

independent coding of spatial layout in the human brain. *Curr Biol, 21*(11), 984-

989. doi: S0960-9822(11)00480-5 [pii]

10.1016/j.cub.2011.04.038

Woodhead, Z. V., Wise, R. J., Sereno, M., & Leech, R. (2011). Dissociation of sensitivity to

spatial frequency in word and face preferential areas of the fusiform gyrus. *Cereb*

Cortex, 21(10), 2307-2312. doi: 10.1093/cercor/bhr008

bhr008 [pii]

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to

understand sensory cortex. *Nat Neurosci, 19*(3), 356-365. doi: 10.1038/nn.4244

nn.4244 [pii]

Figure legends

**Figure 1. Vision beyond object recognition.** This figure illustrates several important functions that are supported by visual perception. These different functions (object recognition, object use, social cognition, navigation) each require a specific way of processing the contents of a scene. For example, perceiving objects with the goal to use these as tools ("object use") may require representing those objects in terms of their associated hand postures. In general, carrying out these diverse functions demands idiosyncratic representations of the kinds of visual stimuli that they typically involve. We argue that the need for these diverse kinds of representation -- that go beyond the need to encode object form in a view-, size- and location-invariant fashion -- offers an account of the presence and properties of numerous category-selective regions in the ventral temporal cortex.

Figure 2. Category selectivity can be dissociated from visual feature processing. a. Response in tool-selective LOTC increased for objects that were perceived as tools after hands-on training to use these objects to perform tool-like tasks (Weisberg et al., 2007). T = Trained objects, NT = Not-trained objects. **b.** Selectivity for animals in ventral temporal cortex was preserved for visually matched object pairs, indicating that this selectivity is not fully driven by shape features (Proklova et al., 2016).

Figure 3. Category-selective activity in congenitally blind individuals. a. Tool-selective LOTC-tool (defined in sighted participants using a standard visual localizer) responds more strongly to aurally presented names of tools (T) than to names of animals (A) and objects (O) (Peelen et al., 2013). **b**. Letter-selective VWFA (defined in sighted participants using a standard visual localizer) responds selectively when blind participants read letters using a sensory-substitution device that converts images to soundscapes (Striem-Amit et al., 2012).
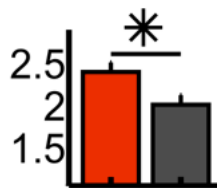
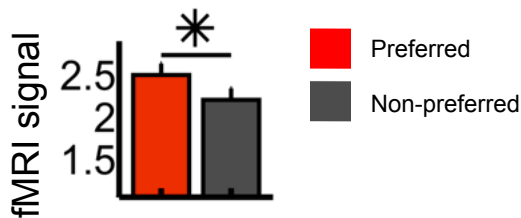object recognition

object use

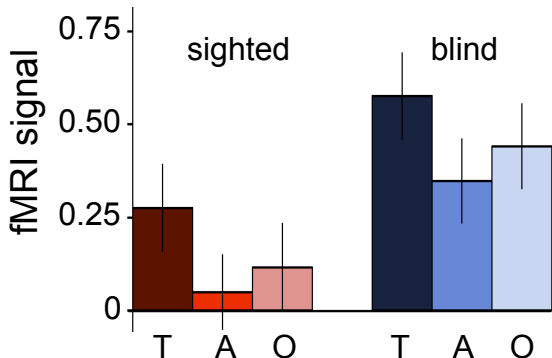social cognition
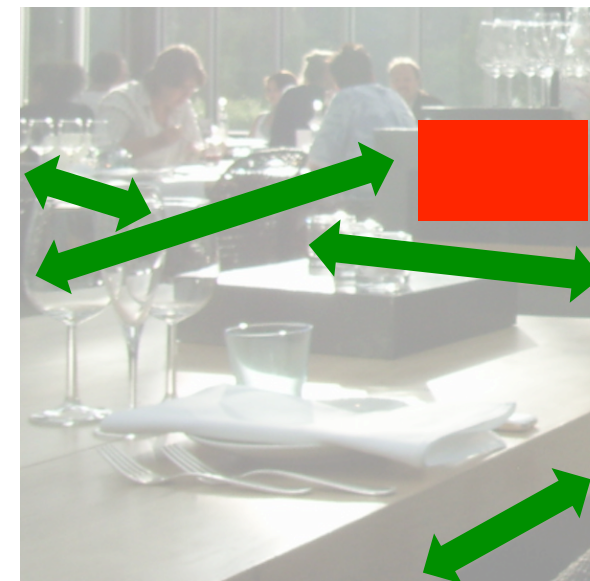
navigation

a. LOTC-tool reflects manipulation experience
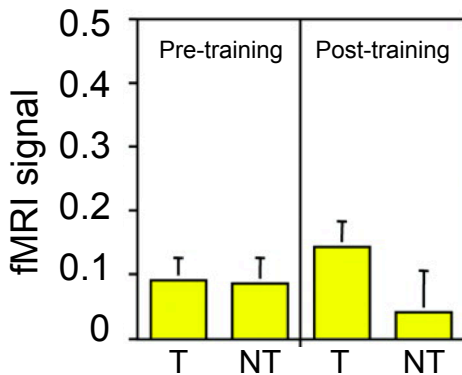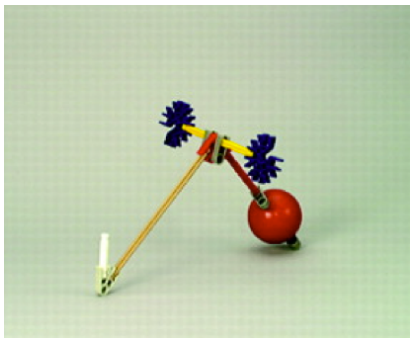
b. Category selectivity for shape-matched objects
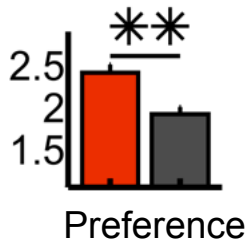
a. LOTC-tool responds to tool names in blind

"spoon" (T)
"snail" (A)
"stool" (O)

b. VWFA responds to letters in blind using SSD